ELSEVIER

Biochimica et Biophysica Acta 1471 (2001) C1–C10

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.elsevier.com/locate/bba

Oncology corner

The use of common genetic polymorphisms to enhance the epidemiologic study of environmental carcinogens

N. Rothman *, S. Wacholder, N.E. Caporaso, M. Garcia-Closas, K. Buetow, J.F. Fraumeni Jr.

Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD 20892, USA

Received 16 May 2000; accepted 15 August 2000

Abstract

Overwhelming evidence indicates that environmental exposures, broadly defined, are responsible for most cancer. There is reason to believe, however, that relatively common polymorphisms in a wide spectrum of genes may modify the effect of these exposures. We discuss the rationale for using common polymorphisms to enhance our understanding of how environmental exposures cause cancer and comment on epidemiologic strategies to assess these effects, including study design, genetic and statistical analysis, and sample size requirements. Special attention is given to sources of potential bias in population studies of gene–environment interactions, including exposure and genotype misclassification and population stratification (i.e., confounding by ethnicity). Nevertheless, by merging epidemiologic and molecular approaches in the twenty-first century, there will be enormous opportunities for unraveling the environmental determinants of cancer. In particular, studies of genetically susceptible subgroups may enable the detection of low levels of risk due to certain common exposures that have eluded traditional epidemiologic methods. Further, by identifying susceptibility genes and their pathways of action, it may be possible to identify previously unsuspected carcinogens. Finally, by gaining a more comprehensive understanding of environmental and genetic risk factors, there should emerge new clinical and public health strategies aimed at preventing and controlling cancer.

Keywords: Cancer; Susceptibility; Environment; Genetics; Gene–environment interaction

1. Cancer and the environment

In 1775, a British surgeon described a high frequency of scrotal cancer among chimney sweeps exposed to coal tar, establishing one of the first links between environmental exposure and cancer [1]. Today, the science of epidemiology – the study of the distribution and determinants of cancer in human populations – continues to be a primary source of clues and knowledge about the causes of cancer. *Descriptive studies* have played an important role in estimating the impact of the environment on cancer etiology, which can be roughly gauged by the international variation in cancer statistics gathered from the volume *Cancer Incidence in Five Continents* [2] (Table 1). The differences between areas with the highest and lowest rates range from 50- to 150-fold for melanoma and for cancers of the nasopharynx, prostate, and liver, to about five-fold for leukemia. Some of the variation can be attributed to diagnostic and reporting practices, as well as to genetic factors for certain tumors (e.g., melanoma, which tends to

* Corresponding author.
E-mail: rothmann@epndce.nci.nih.gov

0304-419X/01/$ – see front matter
PII: S0304-419X(00)00021-4

affect fair-skinned populations). However, the available evidence for most tumors suggests that environmental factors are driving the geographic patterns. Further, these ratios may actually underestimate the true global variation because some regions with exceptionally high rates are not covered by population-based tumor registries.

In their comprehensive report on the avoidable causes of cancer, Doll and Peto [3] took the rates from the lowest-risk countries and subtracted them from the rates prevailing in the USA to produce estimates that about 75–80% of all cancer in the United States is due to environmental factors and is, thus, potentially avoidable. The lowest risk for each cancer was considered to be the baseline or background level for so-called spontaneous tumors that, at least in theory, cannot be prevented. Perhaps the most persuasive evidence for environmental factors is reflected in studies of migrant populations, which generally reveal a shift in cancer risk toward that prevailing in the host country, with changes often becoming apparent among first generation immigrants [4].

Another clue to the influence of the environment in cancer etiology is the temporal variation observed for several cancers, even after accounting for improvements in detection or reporting. Of special interest are those tumors that have shown significant increases in incidence in the Surveillance, Epidemiology and End Results (SEER) program of the National Cancer Institute from 1973 to 1995, and which are still rising based on the latest figures. These tumors include melanoma, pleural mesothelioma, non-Hodgkin's lymphoma, hepatocellular carcinoma, renal cell carcinoma, esophageal adenocarcinoma and female lung cancer [5].

While descriptive studies of cancer have provided an important source of clues to causal factors, the major contribution of epidemiology has been to test etiologic hypotheses by *analytical studies*, notably cohort or case–control designs, that have uncovered the pivotal role of lifestyle and other environmental determinants of cancer. The growing list of environmental exposures linked to cancer includes tobacco smoking, alcohol, diet, ultraviolet and ionizing radiation, occupational toxins, infectious agents, body weight, exercise, reproductive history, and certain medications [6]. There are also many forms of cancer that appear to have a substantial environmental component, but whose etiology remains obscure or incompletely understood. A major challenge now is to further identify and sharpen our understanding of the environmental determinants of specific cancers, even at low levels of exposure. This challenge is becoming increasingly difficult to meet through traditional epidemiologic approaches for several reasons. First, exposures often cannot be directly measured and, in some study designs, we must depend on a person's recall of events that occurred years or decades before. Second, it is more difficult to isolate the effect of individual exposures that tend to occur together. Third, low levels of risk from common exposures are inherently difficult to identify in the general population.

This problem can be alleviated, in part, by more accurate measures of environmental exposures. The approaches include validated questionnaires that are linked to databases of dietary [7], occupational [8–10] and environmental exposures [11–13] and are less prone to reporting bias, as well as direct measurements of exposures where feasible (e.g., air, water and soil contaminants). The collection of blood and urine samples in large cohort studies along with advances in chemical and molecular analytic methods should further increase the accuracy of exposure assessment in epidemiologic studies [14,15].

2. Genetic modifiers of environmental cancer risk

Although increasing our ability to measure cancer-causing exposures is important, it is necessary also to understand cancer susceptibility, particularly if an environmental exposure increases cancer risk primarily in a vulnerable subgroup of the population [16]. For example, several decades after the report of scrotal cancer among chimney sweeps, it was noted that not all chimney sweeps exposed to soot were affected, suggesting that constitutional factors may play a role as well [1]. In many other settings, tumors have arisen in only a fraction of the population exposed for long periods to relatively high levels of an established human carcinogen, such as tobacco smoke. Why?

Although an element of chance is likely to play a role in the complex and multistage stochastic process of carcinogenesis, there is mounting evidence that

Table 1
International variation in cancer incidence

Type of cancer	H/L	Highest rates	Lowest rates
Melanoma	155	Australia	Japan
Nasopharynx	100	Hong Kong	UK
Prostate	70	USA (Black)	China
Liver	50	China	Canada
Cervix uteri	28	Brazil	Israel
Stomach	22	Japan	Kuwait
Lung	19	USA (Black)	India
Colon	19	USA (White)	India
Bladder	16	Switzerland	India
Pancreas	11	USA (Black)	India
Ovary	8	New Zealand (Maori)	Kuwait
Breast	7	Hawaii (Hawaiian)	Israel (non-Jews)
Leukemia	5	Canada	India

genetic factors may influence an individual's susceptibility (or resistance) to cancer. Of special interest is the substantial inter-individual variation in genes whose products metabolize carcinogens and anti-carcinogens, repair DNA damage, and maintain cell cycle control and immune function. These observations have given rise to the hypothesis that the carcinogenic risk of many exogenous and endogenous exposures may be modified by common genetic polymorphisms in one or more of these processes.

Some of the first candidate genes (e.g., NAT2, CYP2D6) to be incorporated into epidemiologic studies were identified in the 1950s and 1960s when patients who developed side effects from certain drugs were found to metabolize these drugs differently [17]. A second source of candidate genes emerged from the study of metabolic gene families (e.g., cytochrome P450s) in animal models and the discovery of polymorphic human homologues [18]. Currently, the number of polymorphic genes that may modify the effects of known or suspected carcinogens is rapidly increasing (Table 2) [18]. Although the study of genetic polymorphisms in cancer etiology is still in its infancy, some promising leads have emerged. For example, a recent meta-analysis suggested that the association between smoking and bladder cancer is most pronounced among individuals with the NAT2 slow acetylation phenotype [19]. Also, there is some evidence that the protection against colorectal cancer provided by dietary folate is enhanced among individuals homozygous for the methylenetetrahydrofolate reductase (MTHFR) C677T allele [20,21].

There are several reasons why incorporating common genetic polymorphisms into epidemiologic studies will enhance our understanding of the relationship between environmental exposures and cancer: (1) by characterizing the effects of established carcinogens among people with particular genetic variants, one can gain mechanistic insights into the origins of cancer; (2) by identifying and studying population subgroups that are genetically susceptible to a particular carcinogen, one can uncover the low levels of risk associated with certain common exposures; and (3) by determining which susceptibility genes are associated with a given cancer, one can generate insights into the potential carcinogens acted upon by these gene products.

Table 2
Environmental exposures that may be modified by cancer susceptibility genes

Exposure	Cancer site	Proposed modifier gene
Aflatoxin	Liver	EPHX1
Alcohol	Esophagus, oral cavity, liver	ADH3, ALDH2, CYP2E1
Aromatic amines	Bladder	NAT1, NAT2
Nitrosamines	Stomach, nasopharynx	CYP2E1
Chlorinated solvents	Kidney	GSTT1
Benzene	Leukemia	CYP2E1, NQO1, MPO
Estrogens	Breast, endometrium	CYP17, CYP19, CYP1B1, COMT1
Androgens	Prostate	AR, SRD5A2
Ionizing radiation	Leukemia	XPD, XPF
Infectious agents	Cervix (HPV), nasopharynx (EBV)	HLA
Polycyclic aromatic hydrocarbons	Lung, larynx	CYP1A1, GSTM1, EPHX1

Several critical issues must be addressed to identify the most effective epidemiologic strategies for evaluating the interplay between common genetic polymorphisms and environmental exposures in determining cancer risk in the general population. These include (1) study design; (2) genetic analysis; (3) statistical analysis; (4) sample size requirements; and (5) potential sources of bias.

3. Study designs for evaluating gene–environment interactions

Although family-based designs have led the way in identifying high-penetrance mutations that confer an exceptional risk of tumors in cancer-prone families, these mutations are generally uncommon and appear to contribute to only a relatively small proportion of all cancers in the general population. Further, family designs generally have limited power to identify common polymorphisms with low penetrance and low relative risk [22,23]. Because these alleles are common, however, the fraction of disease due to a particular polymorphism (i.e., the attributable risk) may be substantial and, thus, have important public health implications.

Caporaso and Goldstein [24] have contrasted the study of uncommon, high-penetrance mutations with common low-penetrance polymorphisms (Table 3). They and several others have concluded that the primary approach to the study of common genetic variants and their potential interactions with common environmental exposures should be through population-based epidemiologic studies [23–28]. Population-based studies have been a successful time-tested approach in detecting the environmental causes of cancer, and are well suited to identify the effects of common polymorphic genes and their interactions with environmental exposures. Further, they have the advantage of providing direct estimates of relative risk, absolute risk (penetrance), and the fraction of disease due to environmental exposures, to genetic variants, and to their interactions.

There are two complementary epidemiologic designs that are used to identify the causes of chronic disease including cancer: the case–control and the cohort study. The relative strengths and weaknesses of each approach for the study of gene–environment interactions have been discussed [14,27,29] and are briefly described below.

3.1. Case–control studies

An important advantage of the case–control design is that a large number of cases with common or uncommon tumors can be enrolled in a relatively short period of time. Further, very detailed assessment of specific exposures can be carried out, and it is possible to over-sample population groups of special interest (e.g., younger cases, minorities, groups with particular exposures) that may be under-represented in cohort studies [27]. Biologically intensive studies can be carried out with this design, particularly in hospital-based studies where more types, larger amounts and more extensive processing of biological samples can take place [27]. Also, case–control studies can be quickly fielded in response to new public health concerns about exposures that are not routinely assessed in cohort studies. Among the important limitations of case–control studies are selection bias that may arise from relatively low participation rates, recall bias and difficulty in assessing past exposures.

A modification of the case–control method has

Table 3
High-penetrance, uncommon mutations versus low-penetrance polymorphisms

Characteristic	Mutation	Polymorphism
Gene frequency	Generally uncommon	Common ($>1\%$)
Penetrance	High	Low
Absolute/relative risk	Low	High
Population attributable risk	Low	High
Role of environment	Modest	Critical
Study setting	Family	Population
Study type	Linkage	Case–control/cohort

been recently developed, where cases only are used to test for the presence of gene–environment interactions [30–32]. As long as the underlying assumption of this approach is met, i.e., that the allele and the exposure are independent, the case-only method can be an efficient approach to testing for interactions.

3.2. Prospective cohort studies

A major advantage of the prospective cohort study is that both interview and biological measures of exposure are generally considered unbiased because they are collected before cancer diagnosis, and it may be possible to collect exposure data at multiple times over the period of follow-up. Further, a cohort study can be used to estimate the risks associated with a given allele for multiple common tumors and other common diseases in the same population. In addition, cohort studies do not suffer from the potential selection bias found in case–control studies. The nested case–control design is the usual method of choice for the study of particular tumors arising in the cohort, since only a relatively small portion of cohort members needs to be included in the analysis. However, cohort studies tend to be costly and require substantial personnel and a long-term infrastructure to maintain and follow up the study population. Also, most cohort studies will not have sufficient numbers of subjects with less common tumors to evaluate risks associated with gene–environment interactions.

More than one million people in the USA, Europe, and Asia will soon be enrolled into cohort studies that are collecting DNA and questionnaire data on dietary and other environmental factors [29]. At the same time, an increasing number of large case–control studies of various forms of cancer are being fielded, with biospecimen collections that include a source of genomic DNA. It is clear that both kinds of studies will be needed to fully investigate the role of environmental and genetic factors and their interactions in cancer etiology.

4. Genetic analysis

The candidate genes evaluated in previous studies of cancer susceptibility have been limited so far to a relatively small number of loci, and almost certainly represent a very small subset of the actual genes with variants that modulate cancer risk. Recent advances in human genomics, however, make it practical to greatly expand the number of genes that might be considered in such studies. For example, the NCI's Cancer Genome Anatomy Project and its related efforts have identified more than 50 000 genes/transcripts and greater than 10 000 variants within these genes [33]. It is anticipated that in the very near term all genes will be cataloged and their common variants identified, although the impact on gene product will be unknown in many instances.

The research opportunity presented by the availability of such information is still constrained by technical limits in the characterization and interpretation of large collections of genetic variants. Of special concern are the large number of individual assays that must be performed; the amount of sample material required from an individual to conduct these assays; the need for a multiplicity of assays on large numbers of individuals; the cost per individual assay; and novel challenges in data management and analysis. Many technological approaches have been proposed to solve these problems, notably massively parallel hybridization assays such as DNA chips [34]. This approach performs thousands to tens of thousands of genotyping assays simultaneously, discriminating genetic variants by differential hybridization of the alternative forms.

Another technical approach extends the concept of sequencing by adding tags that differ depending on an individual's DNA constitution [34]. The assays can be performed in multiplex and very rapidly. There is also significant interest in mass spectrometry approaches, wherein variations in DNA constitution are determined by measuring the differences in mass associated with different sequences [35–38]. Again, these assays can be performed in multiplex with each test requiring only seconds to perform. All these approaches use only small amounts of sample material and hold the promise of low expense, but none has moved beyond the level of proof of principle.

Using current technological capabilities, it is already possible to develop strategies to efficiently evaluate the expanded sets of genetic variants that have

already been identified. A logical approach utilizes existing knowledge of biological pathways to prioritize genes for study. One systematically evaluates variants in collections of genes and gene family members that, based on current knowledge, are suspected to modify the effects of a particular environmental exposure [34]. Traditionally, one analyzes these variants in individual DNA samples from cases and controls. Alternatively, one could evaluate many genetic variants at one time by pooling a relatively small amount of DNA from each case and control. Genetic variants are, thus, assessed in the aggregate, and allele frequencies in each pool are contrasted. Only the subset of alleles showing promising results would be expanded to individual testing and evaluation of interactions with environmental and other genetic risk factors. Ultimately, the results from these approaches will help prioritize which genetic variants need to be studied at the biochemical level to provide biological evidence that supports the association.

5. Statistical analysis

The analysis of exposure and genetic effects should begin, in general, with the assessment of their crude main effects (that is, the effect of the exposure ignoring the gene, and vice versa). The next step in the analysis of environmental and genetic factors is to construct a 2×2 interaction table, as illustrated in Table 4. In this example, the odds ratios (ORs), which are estimates of the relative risk when the disease is rare in the population, are presented with individuals who lack both the exposure and the at-risk genotype as the reference category. The OR is 1.0 for the at-risk genetic variant in the absence of exposure, 2.0 for the exposure without the at-risk allele, and 3.0 for having both factors. These values are similar to the ORs estimated from the interaction of smoking and the NAT2 slow acetylation genotype in the risk of developing bladder cancer [19].

Table 4
Odds ratios for a hypothetical example of a 1.5-fold multiplicative gene–environment interaction

Exposure status	At-risk genotype	
	−	+
Unexposed	1.0	1.0
Exposed	2.0	3.0

The interaction table provides useful information concisely. It provides estimates and corresponding tests of the effects of each factor in the presence and in the absence of the other. The following specific questions can be addressed from this table:

1. Is there an effect of the exposure, as measured by the OR, in either of the genotype strata?
2. Is the effect of exposure in one genetic stratum different from the effect of exposure in the other stratum?
3. If an intervention successfully eliminated exposure in everyone, by what percentage would the cancer rate be reduced in each stratum?
4. Is the potential reduction in cancer rate from an intervention that successfully eliminates exposure greater in one stratum than another?

With the advent of relatively inexpensive high-throughput genotyping, analytic studies will be able to assess both the crude main effects of polymorphisms in thousands of different genes and whether or not they modify the effects of environmental exposures. The current paradigm of statistical testing divides the results into 'significant' and 'not significant'. With adjustment for multiple comparisons using the Bonferroni correction [23], very large studies will be required to have sufficient power to find polymorphisms that may cause a substantial fraction of the tumor under study. Without adjustment for multiple comparisons, the scientific community will be inundated with meaningless 'positive' findings. One way to deal with this problem may be to enhance Bayesian analytical approaches [39] that incorporate a priori knowledge about the action of genetic variants and relevant exposures from previous experimental and epidemiologic studies. In addition, it will be important to develop procedures by which investigators of different studies can quickly report both 'positive' and 'negative' results so that findings can be replicated in a timely fashion, and combined with other studies in a simple, rational and coordinated manner. Finally, improved meta-analytic procedures for providing global assessments of results will be needed as well.

6. Sample size requirements

The first concern with study power usually focuses on having an adequate sample size to detect the crude main effect. For example, the sample size needed to study an environmental exposure or a particular genotype with an OR of 1.5 and prevalence of 10% in the general population would be about 900 cases and 900 controls (assuming 80% power and performing a two-sided test of 0.05). If one is concerned about the problem of multiple comparisons due to testing many independent alleles and plans to carry out a Bonferroni correction for each test, then the sample size would need to be increased. For example, the sample size noted above would need to be doubled if one wished to evaluate and correct for 100 genotypes.

Estimating the sample size needed to test for interactions is a more complex process, which has been commented on extensively [40–46]. It is clear that sample size estimates are highly dependent on assumptions that include the underlying interaction model (e.g., additive or multiplicative), the magnitude of the interaction, the effect of the genetic factor conditioned on the environmental factor and vice versa, the prevalence of the exposure and the polymorphism, and whether exposure is analyzed as a bivariate, polytomous or continuous variable. Further, the need for increasing sample size to take into account multiple comparisons may be necessary as well.

To illustrate, Fig. 1 shows the number of cases (with an equal number of controls) needed to detect a 1.5-fold multiplicative interaction, as the probability of being exposed varies, for different probabilities of having the at-risk genotype. These calculations use the ORs in Table 4, set power equal to 80%, and use an α level for a two-sided test of 0.05. We can see that the smallest required sample size to detect this interaction for common exposures and an at-risk genotype prevalence of 0.50 is 500–600 cases and a similar number of controls. If one is interested in studying less common or more common alleles and exposures, sample size requirements rapidly increase to several thousand cases and controls (Fig. 1). We consider this a fairly realistic scenario, in that the crude main effects of the at-risk genotype and the exposure are modest. For example, at an exposure and genotype prevalence of 0.5, the ORs for the crude main effect of the exposure and the at-risk genotype are 2.5 and 1.3, respectively.

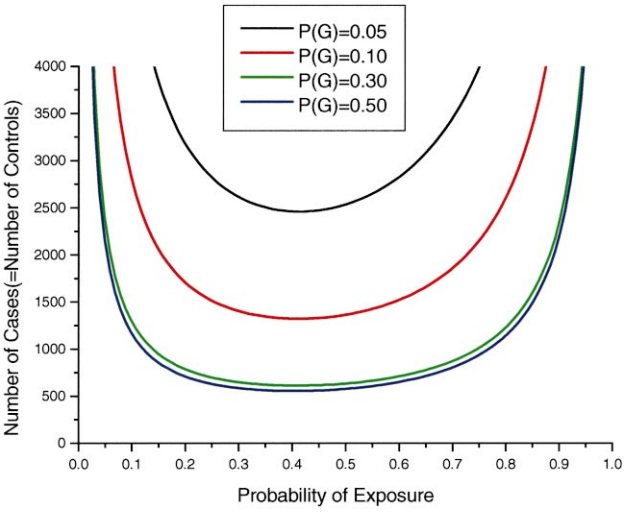

Fig. 1. Number of cases (with an equal number of controls) needed to detect a 1.5-fold multiplicative interaction, as the probability of being exposed varies, for different probabilities of having the at-risk genotype. Power = 80%, two-sided α value = 0.05, and the ORs are 1.0 for the at-risk genetic variant in the absence of exposure, 2.0 for the exposure without the at-risk allele, and 3.0 for having both factors.

Taking into account results from this type of sample size calculation, several research groups are currently enrolling one to several thousand cases and a similar number of controls into case–control studies of particular tumors. Further, as ongoing prospective cohort studies with biological samples mature [29], it will be feasible to carry out large, nested case–control studies of common tumors within these cohorts as well. These studies should have sufficient flexibility to study a wide range of environmental factors and determine if the effects are limited to or enhanced by certain genotypes. We note, however, that when there is interest in studying less common exposures or alleles, sampling schemes may be required to enrich for population subgroups with these risk factors. In the case of less common alleles, alternative study designs that use related controls will generally be needed [25,28,40,47].

7. Sources of potential bias

Potential biases in classic epidemiologic study de-

signs are well-recognized [48], and apply to molecular epidemiology studies as well. Some of the most important biases can be caused by incomplete ascertainment of cases, poor selection of controls, low response rates, and any source of error in the collection and analysis of data that treats cases differently than controls. There are additional potential biases that are particularly relevant to the study of gene–environment interactions. Here, we comment on two key issues: (1) misclassification of exposure and genetic data, and (2) population stratification.

7.1. Exposure and genotype misclassification

The evaluation of environmental and genetic risk factors, independently and together, in cancer etiology requires the accurate measurement of both. Rothman et al. [49,50] explored the impact of genotype misclassification on the genotype risk estimate and Garcia-Closas et al. [51] explored the impact of genotype and exposure misclassification on sample size requirements and the estimated interaction effect [51]. It has been shown that modest exposure assessment errors may result in substantial increases in sample size requirements [51]. This problem is compounded by even small errors in genotype assessment. Given the already large sample sizes required to detect interactions between environmental exposures and genetic polymorphisms in cancer risk, it is critical that special efforts be made to collect highly accurate data on both exposure and genotype.

7.2. Population stratification

A criticism of traditional epidemiologic study designs that use unrelated subjects as controls is that they may lead to confounding due to unrecognized ethnic admixture, known in the genetics field as 'population stratification' [52]. Concerns about this issue have contributed, in part, to the development of several alternative study designs that use unaffected relatives as controls and utilize the transmission disequilibrium test to assess disease–allele associations [23,26,52–54]. Wacholder and colleagues have recently quantified the bias from population stratification under a wide range of conditions and concluded that its impact will be relatively minor in well-conducted epidemiologic studies [27,55]. When important confounding caused by population stratification does occur, it should be responsive to the usual design and analytic approaches employed by epidemiologists, perhaps complemented in the future by genetic markers of ethnicity [26,55,56]. Alternative study designs using unaffected related controls have a role in certain situations [28,34,47,53,54,57–60], but these approaches have potential biases and statistical or economical inefficiencies of their own that need to be evaluated [27,47].

8. Conclusion

The epidemiologic evidence assembled to date indicates that environmental exposures, broadly defined to include lifestyle factors, are responsible for most cancer. There is also reason to believe that relatively common polymorphisms in a broad spectrum of genes may modify the risk imparted by exogenous and endogenous exposures. Although the penetrance of common polymorphisms is likely to be relatively small, the high percent of individuals who carry these alleles suggests that common polymorphisms may, in combination with relevant exposures, contribute to a substantial portion of the cancer burden in the general population. A major challenge of cancer epidemiology in the coming years will be to apply the emerging tools of molecular genetics to help generate a comprehensive understanding of the environmental and genetic determinants of cancer. Epidemiologic studies that measure susceptibility genes should provide opportunities to detect low levels of risk due to certain common exogenous (e.g., diet, pollution) and endogenous (e.g., hormones) exposures, to illuminate pathways of action that may point to previously unsuspected carcinogens, and to detect gene–environment interactions that may give rise to new clinical and public health strategies aimed at preventing and controlling cancer.

References

[1] H.A. Waldron, A brief history of scrotal cancer, Br. J. Ind. Med. 40 (1983) 390–401.
[2] D.M. Parkin, S.L. Whelan, J. Ferlay, L. Raymond, I. Young, Cancer Incidence in Five Continents, Volume VII, IARC Press, Lyon, 1997.

[3] R. Doll, R. Peto, The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today, J. Natl. Cancer Inst. 66 (1981) 1191–1308.
[4] D.B. Thomas, M.R. Karagas, Migrant studies, in: D. Schottenfeld, J.F. Fraumeni, Jr. (Eds.), Cancer Epidemiology and Prevention, Oxford University Press, New York, 1996, pp. 236–54.
[5] L.A.G. Ries, C.I. Kosary, B.F. Hankey, B.A. Miller, A. Harras, B.K. Edwards, SEER Cancer Statistics Review, 1973–1994, National Cancer Institute, Bethesda, MD, 1994.
[6] D. Schottenfeld, J.F. Fraumeni, Jr., Cancer Epidemiology and Prevention, Oxford University Press, New York, 1996.
[7] R. Sinha, N. Rothman, C.P. Salmon, M.G. Knize, E.D. Brown, C.A. Swanson, D. Rhodes, S. Rossi, J.S. Felton, O.A. Levander, Heterocyclic amine content in beef cooked by different methods to varying degrees of doneness and gravy made from meat drippings, Food Chem. Toxicol. 36 (1998) 279–287.
[8] W.F. Stewart, P.A. Stewart, Occupational case–control studies: I. Collecting information on work histories and work-related exposures, Am. J. Ind. Med. 26 (1994) 297–312.
[9] P.A. Stewart, W.F. Stewart, Occupational case–control studies: II. Recommendations for exposure assessment, Am. J. Ind. Med. 26 (1994) 313–326.
[10] P.A. Stewart, W.F. Stewart, E.F. Heineman, M. Dosemeci, M. Linet, P.D. Inskip, A novel approach to data collection in a case–control study of cancer and occupational exposures, Int. J. Epidemiol. 25 (1996) 744–752.
[11] M. Koivusalo, E. Pukkala, T. Vartiainen, J.J. Jaakkola, T. Hakulinen, Drinking water chlorination and cancer – a historical cohort study in Finland, Cancer Causes Control 8 (1997) 192–200.
[12] K.P. Cantor, C.F. Lynch, M.E. Hildesheim, M. Dosemeci, J. Lubin, M. Alavanja, G. Craun, Drinking water source and chlorination byproducts. I. Risk of bladder cancer [see comments], Epidemiology 9 (1998) 21–28.
[13] M.H. Ward, J.R. Nuckols, S.J. Weigel, S.K. Maxwell, K.P. Cantor, R.S. Miller, Identifying populations potentially exposed to agricultural pesticides using remote sensing and a Geographic Information System, Environ. Health Perspect. 108 (2000) 5–12.
[14] N. Rothman, W.F. Stewart, P.A. Schulte, Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories, Cancer Epidemiol. Biomarkers Prev. 4 (1995) 301–311.
[15] W. Willett, Nutritional Epidemiology, Oxford University Press, New York, 1998, pp. 486–487.
[16] N. Rothman, R.B. Hayes, Using biomarkers of genetic susceptibility to enhance the study of cancer etiology, Environ. Health Perspect. 103 (1995) 291–295.
[17] W.W. Weber, Human drug-metabolizing enzyme variants, in: Pharmacogenetics, Oxford University Press, New York, 1997, pp. 131–216.
[18] N. Caporaso, Selection of candidate genes for population studies, IARC Sci. Publ. 148 (1999) 23–36.
[19] P.M. Marcus, R.B. Hayes, P. Vineis, M. Garcia-Closas, N.E. Caporaso, H. Autrup, R.A. Branch, J. Brockmoller, T. Ishizaki, A.E. Karakay, J.M. Ladero, S. Mommsen, H. Okkels, M. Romkes, I. Roots, N. Rothman, The interactive effect of NAT2 acetylation status and cigarette smoking on bladder cancer risks: a case-series meta-analysis. Cancer Epidemiol. Biomarkers Prev. 2000 (in press).
[20] J. Ma, M.J. Stampfer, E. Giovannucci, C. Artigas, D.J. Hunter, C. Fuchs, W.C. Willett, J. Selhub, C.H. Hennekens, R. Rozen, Methylenetetrahydrofolate reductase polymorphism, dietary interactions, and risk of colorectal cancer, Cancer Res. 57 (1997) 1098–1102.
[21] M.L. Slattery, J.D. Potter, W. Samowitz, D. Schaffer, M. Leppert, Methylenetetrahydrofolate reductase, diet, and risk of colon cancer, Cancer Epidemiol. Biomarkers Prev. 8 (1999) 513–518.
[22] D.A. Greenberg, Linkage analysis of necessary disease loci versus susceptibility loci, Am. J. Hum. Genet. 52 (1993) 135–143.
[23] N. Risch, K. Merikangas, The future of genetic studies of complex human diseases, Science 273 (1996) 1516–1517.
[24] N. Caporaso, A. Goldstein, Cancer genes: single and susceptibility: exposing the difference, Pharmacogenetics 5 (1995) 59–63.
[25] Q.H. Yang, M.J. Khoury, Evolving methods in genetic epidemiology. 3. Gene–environment interaction in epidemiologic research, Epidemiol. Rev. 19 (1997) 33–43.
[26] M.J. Khoury, Q.H. Yang, The future of genetic studies of complex human diseases: An epidemiologic perspective, Epidemiology 9 (1998) 350–354.
[27] N. Caporaso, N. Rothman, S. Wacholder, Case–control of common alleles and environmental factors, J. Natl. Cancer Inst. Monogr. 26 (1999) 25–30.
[28] A.S. Whittemore, L.M. Nelson, Study design in genetic epidemiology: theoretical and practical considerations, J. Natl. Cancer Inst. Monogr. 26 (1999) 61–69.
[29] B. Langholz, N. Rothman, S. Wacholder, D.C. Thomas, Cohort studies for characterizing measured genes, J. Natl. Cancer Inst. Monogr. 26 (1999) 39–42.
[30] W.W. Piegorsch, C.R. Weinberg, J.A. Taylor, Non-hierarchical logistic-models and case-only designs for assessing susceptibility in population-based case–control studies, Stat. Med. 13 (1994) 153–162.
[31] M.J. Khoury, W.D. Flanders, Nontraditional epidemiologic approaches in the analysis of gene–environment interaction: Case–control studies with no controls!, Am. J. Epidemiol. 144 (1996) 207–213.
[32] D.M. Umbach, C.R. Weinberg, Designing and analysing case–control studies to exploit independence of genotype and exposure, Stat. Med. 16 (1997) 1731–1743.
[33] R.L. Strausberg, C.A. Dahl, R.D. Klausner, New opportunities for uncovering the molecular basis of cancer, Nature Genet. 15 (1997) 415–416.
[34] D.J. Schaid, K. Buetow, D.E. Weeks, E. Wijsman, S.W. Guo, J. Ott, C. Dahl, Discovery of cancer susceptibility genes: study designs, analytic approaches, and trends in technology, J. Natl. Cancer Inst. Monogr. 26 (1999) 1–16.

[35] A. Braun, D.P. Little, H. Koster, Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry, Clin. Chem. 43 (1997) 1151–1158.
[36] A. Braun, D.P. Little, D. Reuter, B. Muller-Mysok, H. Koster, Improved analysis of microsatellites using mass spectrometry, Genomics 46 (1997) 18–23.
[37] D.P. Little, A. Braun, M.J. O'Donnell, H. Koster, Mass spectrometry from miniaturized arrays for full comparative DNA analysis, Nature Med. 3 (1997) 1413–1416.
[38] D.P. Little, A. Braun, B. Darnhofer-Demar, H. Koster, Identification of apolipoprotein E polymorphisms using temperature cycled primer oligo base extension and mass spectrometry, Eur. J. Clin. Chem. Clin. Biochem. 35 (1997) 545–548.
[39] D. Malakoff, Bayes offers a 'new' way to make sense of numbers, Science 286 (1999) 1460–1464.
[40] A.M. Goldstein, R.T. Falk, J.F. Korczak, J.H. Lubin, Detecting gene–environment interactions using a case–control design, Genet. Epidemiol. 14 (1997) 1085–1089.
[41] A.M. Goldstein, N. Andrieu, Detection of interaction involving identified genes: available study designs, J. Natl. Cancer Inst. Monogr. 26 (1999) 49.
[42] M. Garcia-Closas, J.H. Lubin, Power and sample size calculations in case-control studies of gene–environment interactions: comments on different approaches [comment], Am. J. Epidemiol. 149 (1999) 689–692.
[43] P.G. Smith, N.E. Day, The design of case–control studies: the influence of confounding and interaction effects, Int. J. Epidemiol. 13 (1984) 356–365.
[44] S.J. Hwang, T.H. Beaty, K.Y. Liang, J. Coresh, M.J. Khoury, Minimum sample-size estimation to detect gene environment interaction in case–control designs, Am. J. Epidemiol. 140 (1994) 1029–1037.
[45] J.H. Lubin, M.H. Gail, On power and sample size for studying features of the relative odds of disease, Am. J. Epidemiol. 131 (1990) 552–566.
[46] I. Foppa, D. Spiegelman, Power and sample size calculations for case-control studies of gene–environment interactions with a polytomous exposure variable, Am. J. Epidemiol. 146 (1997) 596–604.
[47] J.S. Witte, W.J. Gauderman, D.C. Thomas, Asymptotic bias and efficiency in case–control studies of candidate genes and gene–environment interactions: basic family designs, Am. J. Epidemiol. 149 (1999) 693–705.
[48] K.J. Rothman, S. Greenland, Modern Epidemiology, Lippincott-Raven, Philadelphia, PA, 1998.
[49] N. Rothman, W.F. Stewart, N.E. Caporaso, R.B. Hayes, Misclassification of genetic susceptibility biomarkers: implications for case–control studies and cross-population comparisons, Cancer Epidemiol. Biomarkers Prev. 2 (1993) 299–303.
[50] N. Rothman, M. Garcia-Closas, W.T. Stewart, J. Lubin, The impact of misclassification in case–control studies of gene–environment interactions, IARC Sci. Publ. 148 (1999) 89–96.
[51] M. Garcia-Closas, N. Rothman, J. Lubin, Misclassification in case–control studies of gene–environment interactions: assessment of bias and sample size, Cancer Epidemiol. Biomarkers Prev. 8 (1999) 1043–1050.
[52] E.S. Lander, N.J. Schork, Genetic dissection of complex traits, Science 265 (1994) 2037–2048.
[53] R.S. Spielman, W.J. Ewens, Transmission disequilibrium test (TDT) for linkage and linkage disequilibrium between disease and marker, Am. J. Hum. Genet. 53 (1993) 863.
[54] R.S. Spielman, W.J. Ewens, The TDT and other family-based tests for linkage disequilibrium and association, Am. J. Hum. Genet. 59 (1996) 983–989.
[55] S. Wacholder, N. Rothman, N. Caporaso, Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias, J. Natl. Cancer Inst. 92 (2000) 1151–1158.
[56] J.K. Pritchard, N.A. Rosenberg, Use of unlinked genetic markers to detect population stratification in association studies, Am. J. Hum. Genet. 65 (1999) 220–228.
[57] D.J. Schaid, S.J. Jacobsen, Biased tests of association: Comparisons of allele frequencies when departing from Hardy–Weinberg proportions, Am. J. Epidemiol. 149 (1999) 706–711.
[58] D.J. Schaid, C. Rowland, Use of parents, sibs and unrelated controls for detection of associations between genetic markers and disease, Am. J. Hum. Genet. 63 (1998) 1492–1506.
[59] D.C. Thomas, Design of gene characterization studies: an overview, J. Natl. Cancer Inst. Monogr. 26 (1999) 17–23.
[60] W.J. Gauderman, J.S. Witte, D.C. Thomas, Family-based association studies, J. Natl. Cancer Inst. Monogr. 26 (1999) 31–37.